ICMA Centre

Henley Business School
UNIVERSITY OF READING

Discussion Paper

Do forecasters target first or later releases of national accounts data?

June 2017

Michael P Clements
ICMA Centre, Henley Business School, University of Reading

Discussion Paper Number:
ICM-2017-03

The aim of this discussion paper series is to disseminate new research of academic distinction. Papers are preliminary drafts, circulated to stimulate discussion and critical comment. Henley Business School is triple accredited and home to over 100 academic faculty, who undertake research in a wide range of fields from ethics and finance to international business and marketing.

admin@icmacentre.ac.uk

www.icmacentre.ac.uk

Do forecasters target first or later releases of national accounts data?

Michael P. Clements*
ICMA Centre,
Henley Business School,
University of Reading,
Reading RG6 6BA

June 28, 2017

Abstract

We consider whether there is any evidence that macro forecasters attempt to forecast data-vintages beyond the first estimates. Our approach requires that both the first and subsequent estimates are predictable prior to the first estimate being released using publically available information. There is some weak evidence that consensus forecasts of consumers' expenditure target vintage estimates after the first estimate, and that around a fifth of individual forecasters put some weight on later estimates of the macro variables.

*Michael Clements is Professor of Econometrics at the ICMA Centre, Henley Business School, University of Reading, and a member of the Institute for New Economic Thinking at the Oxford Martin School, University of Oxford. Email: m.p.clements@reading.ac.uk. The computations were undertaken using code written in Gauss 14, Aptech Systems.

1 Introduction

Most macro data are subject to revision over time, with different vintage estimates (or data maturities) being published as more complete information becomes available. Which vintage of data should be taken as the 'actual values' for calculating forecast errors in forecast evaluation exercises? In the literature, some studies use first-release actuals, or actuals available after a relatively small number of monthly revisions, or the vintage available immediately prior to a benchmark revision, or even (what might be described as) fully-revised data.[1] In this paper we suggest a simple method that can be used in principle to determine the target, although we show that what is required is that the future revision be predictable before the first estimate of the data value is released, using publically-available information. We show that for four of the five macro-variables we analyze, there is evidence that on average around a fifth of the survey respondents look beyond the first or advance estimate and put some weight on later estimates.

In addition to the empirical evidence we present, a contribution of the paper is to bring to the fore the reasons why it is difficult to determine the vintage of data being targeted. These turn on the degree of predictability of data revisions, and the problem of weak instruments in our testing strategy. To the best of our knowledge, these issues have been largely overlooked in the literature to data. The problems caused by weak instruments can be countered to some extent by partially robust methods, and we briefly discuss relevant methods.

One might ask why it matters. If one were able to reliably determine the vintage of data being targeted, at the level of the individual forecaster, there would be the potential for an improved understanding of various aspects of the expectations process, and in particular, for the study of why forecasters disagree. This is a much studied question, [2] with recent explanations stressing the role of informational rigidities (IR). Under full-information rational expectations (FIRE), in which all agents know the true structure of the economy, and have access to the same information set, agents have identical expectations. IR have been used to explain the empirical finding of disagreement without jettisoning the basic notion that forecasters form their expectations rationally, subject to the information constraints they face. The two leading

[1] Revisions to some macro variables can be substantial. Croushore (2011, p.249, Figure 9.1) illustrates the magnitude of the revisions that have occurred to the growth rate of real residential investment by plotting the initial release of the 1976:Q3 growth rate and all subsequent estimates (up to those made in 2009). The estimates of the annual rate rose from less than 3% to nearly 16%, before ending up at around −5%!

[2] See, for example, Zarnowitz and Lambros (1987), Bomberger (1996), Rich and Butler (1998), Capistrán and Timmermann (2009), Lahiri and Sheng (2008), Rich and Tracy (2010) and Patton and Timmermann (2010), *inter alia*.

contenders are sticky information,[3] and noisy information.[4] However, if it were the case that different individual forecasters target different vintages, then a proportion of observed disagreement may not reflect true disagreement at all but may arise simply because different forecasters are forecasting different things, viz. the vintage of data of the macro variable. Our findings suggest heterogeneity regarding targets is perhaps unlikely to be a major part of the story, at least for real GDP growth. For GDP growth we find little evidence against the null - that the first estimate is being targeted. More generally, of course, forecast-accuracy comparisons among individual forecasters, and assessments of forecast efficiency (e.g., using the approach popularised by Mincer and Zarnowitz (1969)), may depend on the vintage of actual values that is chosen.

The plan of the remainder of the paper is as follows. In section 2 we present a simplified statistical framework to illustrate the prediction problem when data are subject to revision, and show that a simple regression can be used to determine the vintage of data being forecast. Section 3 presents the empirical analysis, and section 4 some concluding remarks.

2 Statistical Framework and Test

We set out a simple statistical framework to address the question of the circumstances under which we can determine the forecaster's target. Suppose that there are just two vintages, and that the first-vintage estimate of y_t, denoted y_t^{t+1}, is a noisy observation on the true value for y_t, denoted by y_t^{t+2}, that is:

$$y_t^{t+1} = y_t^{t+2} + \eta_t \tag{1}$$

where $\eta_t \perp y_t^{t+2}$, and y_t^{t+2} follows an AR(1), say:

$$y_t^{t+2} = f y_{t-1}^{t+1} + v_t. \tag{2}$$

Finally, we assume that data revisions are serially correlated (as in, e.g., Howrey (1978)):

$$\eta_t = h\eta_{t-1} + w_t. \tag{3}$$

We assume a one-period delay, in that the $t+1$-data vintage (given by the superscript) will contain data up to reference period t (given by the subscript). Hence at time t, the information

[3] See *inter alia* Mankiw and Reis (2002) and Mankiw, Reis and Wolfers (2003), and Coibion and Gorodnichenko (2012, 2015).

[4] See Woodford (2002), Sims (2003) and Coibion and Gorodnichenko (2012, 2015), *inter alia*.

set will consist of $\mathcal{I}^t = \left\{y_{t-1}^t, y_{t-2}^t, \ldots; y_{t-2}^{t-1}, y_{t-2}^{t-1}, \ldots;\right\}$.

Statistical frameworks allowing for multiple estimates, which can be news or noise (in the sense of Mankiw and Shapiro (1986), as discussed below), and more general models etc., are provided by Jacobs and van Norden (2011) and Kishor and Koenig (2012), *inter alia*, although (1) to (3) suffice to illustrate the key points.

Consider forecasts of the first and second estimates of reference quarter y_t given information $\mathcal{I}^t$. From (1) we have:

$$E\left(y_t^{t+1}|\mathcal{I}^t\right) = E\left(y_t^{t+2}|\mathcal{I}^t\right) + E\left(\eta_t|\mathcal{I}^t\right).$$

It follows immediately that unless revisions are predictable, i.e., $h \neq 0$ in our framework, then $E\left(\eta_t|\mathcal{I}^t\right) = E\left(w_t|\mathcal{I}^t\right) = 0$ and the prediction of the revision is zero. If the revision is unpredictable, then the forecasts of the different vintage estimates are identical, as shown, and there is no way of knowing whether the forecaster's intention is to forecast the first or second estimate (say), from the data evidence alone.

To make progress, suppose revisions exhibit some degree of predictability, so that $h \neq 0$. Given a reported forecast of y_t, denoted f_t, we wish to determine whether f_t is a forecast of y_t^{t+2}, or of y_t^{t+1}. The intuition underlying our testing strategy is simply that the forecast error calculated using y_t^{t+1} will be related to the expected data revision if the forecaster is targeting the second estimate, but otherwise the forecast error and expected data revision should be uncorrelated. In a regression context, we would expect $\beta = 0$ in (4) when the forecaster is targeting the first estimate (as assumed by the construction of the dependent variable). Of course the future revision is an endogenous explanatory variable, and (4) will need to be estimated by TSLS, with the explanatory variable replaced by the predicted value from the first-stage regression on suitable instruments.

$$y_t^{t+1} - f_t = \alpha + \beta\left(y_t^{t+2} - y_t^{t+1}\right) + \varepsilon_t \tag{4}$$

This explains the requirement that future revisions should be predictable in advance using publically-available information. Given the statistical setup in (1) to (3), obvious instruments are $Z^t = \left\{y_{t-1}^t, y_{t-2}^t, y_{t-2}^{t-1}\right\}$, for example, which includes the revision for y_{t-2}, i.e., $y_{t-2}^t - y_{t-2}^{t-1}$. This follows from the assumption that revisions are serially correlated as in Howrey (1978), and as in the 'spillovers' of Jacobs and van Norden (2011). In practice, any variables which predict the future revision to observation quarter t, at time t, could be used.

In Appendix A we detail the derivation of the population values of β in (4) when the forecaster is targeting the first release, or the second release, when, as shown in (4), the forecast error is constructed using the first-estimate actual value, y_t^{t+1}. We show that the TSLS estimate of β is $\beta = 0$ when the forecaster targets the first estimate, but that $\beta < 0$ when the second estimate is targeted, and that $\beta \to -1$ as $E\left(w_t^2\right) \to 0$, i.e., as the degree of predictability of future revisions increases.

Notice also that we have framed the problem in terms of the first estimate being targeted - the left-hand-side of (4) is the forecast error using the first estimate. And we then consider whether there is evidence that weight is attached to a later release. However this 'normalization' of the forecast error does not mean that y_t^{t+1} ad y_t^{t+2} are being treated differently. Estimation of (4) using Z^t as instruments is equivalent to estimation based on the moments:

$$E\left[(1-\gamma)\, y_t^{t+1} + \gamma y_t^{t+2} - f_t - \alpha \mid Z_t\right] = 0 \tag{5}$$

where $\gamma = -\beta$. From (5) it is evident that we are estimating the weights $(1-\gamma)$:γ on the first and second estimates, such that $\beta = 0$ in (4) implies no weight on the later estimate, and $\beta = -1$ ($\gamma = 1$) implies only the later estimate is targeted.

It is worth noting that the requirement that $y_t^{t+2} - y_t^{t+1}$ is predictable from Z^t is more demanding than that data revisions are noise. The standard characterization of revisions as news or noise by Mankiw and Shapiro (1986) suggests that revisions are predictable if they are noise. Data revisions are news when they add *new* information, and noise when they reduce measurement error. If data revisions are noise, they are unrelated to the later estimate, so that $\gamma_{no} = 0$ in:

$$y_t^{t+2} - y_t^{t+1} = \alpha + \gamma_{no} y_t^{t+2} + \omega_t \tag{6}$$

but are predictable from y_t^{t+1}. For news revisions, $\gamma_{ne} = 0$ in:

$$y_t^{t+2} - y_t^{t+1} = \alpha + \gamma_{ne} y_t^{t+1} + \omega_t \tag{7}$$

that is, revisions are unpredictable given the first estimate. Revisions being noise is not sufficient to ensure predictability in the first-stage regression, because the first-stage regression considers an information set that excludes y_t^{t+1}. Hence our key requirement is that the revision is predictable from an information set that does not include the first estimate.

The empirical findings in the literature regarding news/noise are somewhat mixed. Mankiw and Shapiro (1986) and Faust, Rogers and Wright (2005) provide empirical evidence that data revisions to US real GDP are largely news, while Aruoba (2008) and Corradi, Fernandez and Swanson (2009) provide extensions and more nuanced findings. Clements and Galvão (2017) consider the use of more general information sets than the variable being forecast, but always assume the earlier estimate is in the information set.

The importance of strong instruments for reliable TSLS inference is evident from the recent literature on weak instruments surveyed by Stock, Wright and Yogo (2002). Their rule-of-thumb suggests a first-stage F statistic in excess of 10 for tolerable TSLS inference (see Stock *et al.* (2002), Table 1 p.522 and their discussion for the meaning of 'tolerable'). Anticipating our empirical findings, in the event that the first-stage regressions F statistics suggest TSLS will be unreliable, we will instead report results using two of the partially robust (to weak instruments) estimators discussed by Stock *et al.* (2002): Limited-Information Maximum Likelihood (LIML) and the Fuller-k estimator (Fuller (1977)), both of which can be seen as k-class estimators (see, e.g., Davidson and MacKinnon (1993, ch. 18)). Letting $\gamma = (\alpha, \beta)'$, the general k-class estimator of γ in (4) is:

$$\hat{\gamma}(k) = \left[X'(I - kM_Z)X\right]^{-1}\left[X'(I - kM_Z)y\right]$$

where X is the $T \times 2$ matrix of the right-hand side variables (the first column is a vector of 1's, denoted $\mathbf{i}$, and the second is the endogenous variable denoted Y_1,with typical element $y_t^{t+2} - y_t^{t+1}$), and M_Z is the annihilation matrix formed from the instruments (plus the constant), i.e., $M_Z = I - P_Z$ where P_Z projects onto the space spanned by the instruments. y is the vector with typical element $y_t^{t+1} - f_t$. $k = 0$ delivers OLS, and $k = 1$ TSLS. For LIML, k is given by the smallest root to the determinantal equation $|Y'M_1Y - kY'M_ZY| = 0$, where $Y = (y\ Y_1)$, and M_1 is the annihilation matrix formed from the columns of the non-endogenous right-hand-side variables - here, just a constant. $k = 1$ when there is a single instrument, so that under just-identification LIML is the same as TSLS. When the model is over-identified, $k > 1$.

The Fuller-k estimator sets $k = k_{LIML} - b/(T - K)$, where K is the number of instruments. We implement this with $b = 1$. Stock *et al.* (2002) discuss the merits of these and other estimators when instruments are weak.

The covariance matrix of the parameter estimates is given by $\hat{\sigma}^2 \left[X'(I - kM_Z)X\right]^{-1}$.

In our empirical work we will confine our attention to Z^t as given above. This leaves open the possibility that better instruments may exist for the revisions to some of the variables we consider, so that TSLS of (4) would give rise to reliable inference of β.

3 Empirical Findings

We use the US Survey of Professional Forecasters (SPF) as our source of expectations, as it is the oldest quarterly survey of macroeconomic forecasts in the US.[5] The SPF has been used extensively in research: see the academic bibliography maintained at https://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/academic-bibliography. We consider the forecasts of five variables: real GDP growth (RGDP), real personal consumption expenditures (RCONSUM), real non-residential fixed investment (RNRESIN), real residential fixed investment (RRESIN), and the GDP deflator inflation (PGDP).

We focus on the current-quarter forecasts, e.g., the forecast made in 2010:Q1 (around the middle of the middle month of the quarter, 15th February) of the quarterly growth rate in the first quarter over the fourth quarter of 2009, and so on. The SPF also collects forecasts of the next quarter, and of the quarters up to one year ahead, as well as forecasts of the current and next year's values of the variables. We chose the current quarter forecasts, as the survey forecasters are naturally better able to forecast at the shortest horizon (see, e.g., Clements (2015)) and we suspect are better able to forecast the revised values at this horizon, than at longer horizons, should they choose to do so. But of course the approach could be adapted to consider longer horizon forecasts. We use the forecasts made in response to the 129 surveys from 1981:Q3 to 2013:Q3, inclusive.

We use in conjunction real-time vintage data from the Real Time Data Set for Macroeconomists (RTDSM), also maintained by the Federal Reserve Bank of Philadelphia (see Croushore and Stark (2001)).

The SPF and RTDSM mnemonics are recorded in table 1.

We begin with a consideration of the predictability of future revisions, given the importance of strong instruments in the first-stage of the estimation of (4). Our instruments consist of $Z^t = \left\{y^t_{t-1}, y^t_{t-2}, y^{t-1}_{t-2}\right\}$. The results are shown in table 2.

Our primary focus is on the first and second quarterly estimates. The first quarterly estimate is the advance estimate, which is released roughly one month after the reference quarter. The second quarterly estimate is the vintage available at the middle of quarter two quarters after the reference quarter.[6] To illustrate the timing, consider the forecaster responding to the 2010:Q1 survey. The question we ask is whether her forecast is of the advance value of y in 2010:Q1

[5] The SPF began as the NBER-ASA survey in 1968:4 and runs to the present day. Since June 1990 it has been run by the Philadelphia Fed, renamed as the Survey of Professional Forecasters (SPF): see Zarnowitz (1969) and Croushore (1993).

[6] See e.g., Landefeld, Seskin and Fraumeni (2008) and Fixler, Greenaway-McGrevy and Grimm (2014) for a discussion of the revisions to US national accounts data.

Table 1: Description of Macroeconomic Variables

Variable	SPF code	RTDSM
Real GDP (GNP)	RGDP	ROUTPUT
Real personal consumption	RCONSUM	RCON
Real nonresidential fixed investment	RNRESIN	RINVBF
Real residential fixed investment	RRESINV	RINVRESID
GDP price index	PGDP	P
(implicit deflator, GNP / GDP deflator)		

The SPF data are from the Philadelphia Fed, http://www.phil.frb.org/econ/spf/.
Data from the RTDSM are from http://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/.

which will be released at the end of April, or by the middle of August? When we consider the predictability of the future revision for the first and second estimates, the first stage regression is of $y_t^{t+2} - y_t^{t+1}$ on a constant and $Z^t = \left\{y_{t-1}^t, y_{t-2}^t, y_{t-2}^{t-1}\right\}$. According to the advice in Stock *et al.* (2002), the F statistic is not large enough to suggest the subsequent second-stage of TSLS will yield reliable inference for any of the 5 variables.

It may be that our forecaster is in fact targeting a later vintage estimate than the second quarterly estimate, and to cover this possibility we allow the more mature estimate to be that available 5 quarters after the reference quarter, and fifteen quarters after the reference quarter. As a preliminary, we asses whether these longer-term revisions - the revision between the first and fifth, and between the first and fifteenth estimates - are more or less predictable in advance, and these results are also reported in table 2. For these longer-term revisions, the requirement that the instruments are known at time t necessarily means that the more mature data necessarily relates to a more distant reference quarter For example, the first-stage regressions are, respectively, of $y_t^{t+5} - y_t^{t+1}$ on $y_{t-1}^t, y_{t-5}^t, y_{t-5}^{t-4}$, and of $y_t^{t+15} - y_t^{t+1}$ on $y_{t-1}^t, y_{t-15}^t, y_{t-15}^{t-14}$. The table shows that the first-stage regression F statistics do not satisfy the 'rule of thumb' of an F statistic greater than 10 for these longer-term revisions, either. Generally the F statistics do not reject at the less stringent conventional significance levels either, although for RRESIN there is evidence that the 1st to 5th revision is predictable - using conventional levels, whereas the 1st to 2nd is not.

As a result of the findings for the first-stage regressions in table 2, we next report results using the two estimators described in section 2.

Table 2: First-stage regression F statistics: the Predictability of Data Revisions

	1st and 2nd	1st and 5th	1st and 15th
RGDP	1.80	1.71	1.16
	(0.15)	(0.17)	(0.33)
RCONSUM	6.93	2.96	3.52
	(0.00)	(0.04)	(0.02)
RNRESIN	1.77	1.91	0.96
	(0.16)	(0.13)	(0.42)
RRESIN	1.00	3.09	0.84
	(0.40)	(0.03)	(0.47)
PGDP	0.61	1.11	0.38
	(0.61)	(0.35)	(0.77)

The instrument sets are $\{y_{t-1}^t, y_{t-2}^t, y_{t-2}^{t-1}\}$, for the 1st and 2nd; $\{y_{t-1}^t, y_{t-5}^t, y_{t-5}^{t-4}\}$ for the 1st and 5th; and $\{y_{t-1}^t, y_{t-15}^t, y_{t-15}^{t-14}\}$ for the 1st and 15th. Figures in parentheses are the p-values of the first-stage F statistics.

3.1 The median forecasts

Table 3 records the results of estimating (4) for the median forecasts, using LIML and Fuller-k. We do not reject the null hypothesis that the weight the median forecaster applies to y_t^{t+2} is zero, that is, that $\beta = 0$. Since $\beta = -1$ implies the second estimate is being targeted (and the weight on the first estimate is zero), of interest is a one-sided test of H_0:$\beta = 0$ versus H_1:$\beta < 0$. Although we do not reject the null at conventional significance levels, we do reject the null hypothesis at the 11% level for both LIML and Fuller-k for RCONSUM.[7]

We also calculated results using the 1st and 5th quarterly vintage estimates. For RCONSUM we obtained p-values of 0.98 using both LIML and Fuller-k, thus formally rejecting the null in favour of some weight being put on a later vintage estimate (here, the 5th quarterly vintage estimate). For the other variables the results were qualitatively unchanged for the median forecasts (and are not reported).

3.2 The individual forecasts

The advantage of using the median forecasts is that we have a sample of $N = 129$ forecasts, although our results are uninformative about heterogeneity across forecasters in terms of vintages being targeted, and it may be that the results for the aggregate mask evidence that some

[7] The reported p-value of 0.89 is the probability of obtaining a larger t-statistic than the calculated value under the null hypothesis.

Table 3: Median Forecasts. Inference on β in Equation (4), using the 1st and 2nd quarterly vintage estimates.

SPF variable	N	LIML		Fuller-k	
		$\hat{\beta}$	p-value	$\hat{\beta}$	p-value
RGDP	124	0.21	0.42	0.24	0.40
RCONSUM	124	-0.87	0.89	-0.87	0.89
RNRESIN	124	-0.07	0.53	-0.10	0.54
RRESIN	124	-5.77	0.77	-2.44	0.83
PGDP	124	4.58	0.22	1.88	0.21

individuals do target later estimates. Hence we also estimate (4) separately for each individual who made 50 or more forecasts over the sample period. Table 4 reports the average estimates of $\hat{\beta}$ across individuals for each variable, and the proportion of forecasters for whom we failed to reject $\beta = 0$ in a two-sided test at the 10% level, and in a one-sided test (against $\beta < 0$) at the 10% level.

The individual-level results indicate that, for all the variables other than RGDP, we reject the null that $\beta = 0$ in the direction of putting some weight on the second estimate y_t^{t+2} for between a 14% and 30% of all respondents (or some 3 to 7 individuals).

Finally, we consider whether individuals who put weight on the later vintage (here, the 2nd estimate) for one variable are more likely to do so for other variables. That is, are the (one-sided) rejections of the null reported in table 4 due to the same small number of individuals, or are they randomly spread over respondents? We restrict our attention to the individual forecasters who filed 50 or more forecasts of each of the 5 variables we consider. Using either LIML or Fuller-k we obtained similar results: for the former we found that the one-sided null was rejected for two of the five variables for 4 individuals, and for three 3 of the five variables for 3 individuals. For Fuller-k we rejected for 2 variables for 5 individuals, and for 3 variables for 3 individuals. Although the sample of forecast respondents is small, the number of individuals for whom we observe rejections for more than one variable strongly suggests a pattern: some respondents are more likely than others to target the later estimate when we consider their behaviour across all variables.[8]

[8] As a simple illustration, suppose the probability of rejection (in one-sided test, at the 10% level) is 0.2 (loosely based on the results in table 4 across all 5 variables). Then if rejections across variables are independent for a given forecaster and across forecasters, then the probability that an individual rejects for two variables is 0.04. Then the number of individuals for whom we reject for 2 variables is Binomial (20, 0.04), and the probability of 3 'successes' (i.e., three individuals) is 0.0365, and for rejecting for 3 variables is Binomial(20, 0.008), with probability of 3 successes only 0.0005. Hence the empirical findings are highly unlikely under the assumption

Table 4: Indvividual Respondents. Average results of inference on β in Equation (4) using 1st and 2nd quarterly vintage estimates.

	1 or 2-sided	# forecasters	LIML	Fuller-k
RGDP	2	25	0.96	0.96
	1		0.96	0.96
RCONSUM	2	24	0.92	0.92
	1		0.70	0.70
RNRESIN	2	23	0.96	0.96
	1		0.869	0.86
RRESIN	2	23	0.91	0.91
	1		0.86	0.86
PGDP	2	22	1.00	0.91
	1		0.67	0.67

The table reports the proportion of forecasters for whom the null that H_0: $\alpha = 0$ was not rejected against the alternative that H_0: $\alpha \neq 0$ in a two-sided test at the 10% level ("1 or 2-sided" column equals "2"), and the the proportion for whom the null that H_0: $\alpha = 0$ was not rejected against the one-sided alternative that H_0: $\alpha < 0$ at the 10% level ("1 or 2-sided" column equals "1").

4 Conclusions

Questions concerning the appropriate data vintage of actual values to use in analyses of forecast performance are somewhat vexing. Recent papers typically use an 'early estimate', such as the advance estimate, or the first-quarterly revised estimate (corresponding to the third-monthly estimate), although other choices have been made. The most mature data available - the vintage available at the time the study is undertaken - is less attractive because the estimates are likely to embody the effects of methodological changes and other revisions which could not have been foreseen at the time. Nevertheless, even the revisions between the advance and first-quarterly revised estimates can be substantial.

If these revisions were unforecastable for each survey respondent, then the forecast would be the same whichever vintage the forecaster was putatively targeting, and the question we ask in this paper would be unanswerable. There is some evidence that revisions are predictable using public information, most notably for consumption expenditure, although it appears that the instruments are weak and will result in unreliable inference in our tests of the weights attributed to early and later estimates by forecasters. For this reason we use estimators which are partially robust to weak instruments, and find evidence that some forecasters do not exclusively focus

that rejections are independent across variables.

on the first (advance estimate). For consumption, we reject the null at the 10% level for nearly a third of forecasters, in the direction of some weight being placed on the later vintage estimate (i.e., in favor of $\gamma > 0$ in (5), or equivalently, $\beta < 0$ in (4)).

There are a number of interesting avenues to explore. We have used as instruments those variables directly suggested by our simple model of the revisions process, although other variables may be relevant, and might alleviate the weak instruments problem.

References

Aruoba, S. B. (2008). Data revisions are not well-behaved. *Journal of Money, Credit and Banking*, **40**, 319–340.

Bomberger, W. A. (1996). Disagreement as a measure of uncertainty. *Journal of Money, Credit and Banking*, **28**, 381–392.

Capistrán, C., and Timmermann, A. (2009). Disagreement and biases in inflation expectations. *Journal of Money, Credit and Banking*, **41**, 365–396.

Clements, M. P. (2015). Are professional macroeconomic forecasters able to do better than forecasting trends?. *Journal of Money, Credit and Banking*, **47,2-3**, 349–381. DOI: 10.1111/jmcb.12179.

Clements, M. P., and Galvão, A. B. (2017). Predicting Early Data Revisions to US GDP and the Effects of Releases on Equity Markets. *Journal of Business and Economic Statistics*, **35:3**, 389 – 406.

Coibion, O., and Gorodnichenko, Y. (2012). What can survey forecasts tell us about information rigidities?. *Journal of Political Economy*, *120*(1), 116 – 159.

Coibion, O., and Gorodnichenko, Y. (2015). Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts. *American Economic Review*, *105*(8), 2644–78.

Corradi, V., Fernandez, A., and Swanson, N. R. (2009). Information in the revision process of real-time datasets. *Journal of Business and Economic Statistics*, **27**, 455–467.

Croushore, D. (1993). Introducing: The Survey of Professional Forecasters. *Federal Reserve Bank of Philadelphia Business Review, November*, 3–15.

Croushore, D. (2011). Forecasting with real-time data vintages, chapter 9. In Clements, M. P., and Hendry, D. F. (eds.), *The Oxford Handbook of Economic Forecasting*, pp. 247–267: Oxford University Press.

Croushore, D., and Stark, T. (2001). A real-time data set for macroeconomists. *Journal of Econometrics*, **105(1)**, 111–130.

Davidson, R., and MacKinnon, J. G. (1993). Oxford: University of Oxford.

Faust, J., Rogers, J. H., and Wright, J. H. (2005). News and noise in G-7 GDP announcements. *Journal of Money, Credit and Banking*, **37 (3)**, 403–417.

Fixler, D. J., Greenaway-McGrevy, R., and Grimm, B. T. (2014). The revisions to GDP, GDI, and their major components. *Survey of Current Business*, **August**, 1–23.

Fuller, W. (1977). Some properties of a modification of the limited information estimator. *Econometrica*, **45**, 939–953.

Howrey, E. P. (1978). The use of preliminary data in economic forecasting. *The Review of Economics and Statistics*, **60**, 193–201.

Jacobs, J. P. A. M., and van Norden, S. (2011). Modeling data revisions: Measurement error and dynamics of 'true' values. *Journal of Econometrics*, **161**, 101–109.

Kishor, N. K., and Koenig, E. F. (2012). VAR estimation and forecasting when data are subject to revision. *Journal of Business and Economic Statistics*, **30(2)**, 181–190.

Lahiri, K., and Sheng, X. (2008). Evolution of forecast disagreement in a Bayesian learning model. *Journal of Econometrics*, **144(2)**, 325–340.

Landefeld, J. S., Seskin, E. P., and Fraumeni, B. M. (2008). Taking the pulse of the economy. *Journal of Economic Perspectives*, **22**, 193–216.

Mankiw, N. G., and Reis, R. (2002). Sticky information versus sticky prices: a proposal to replace the New Keynesian Phillips Curve. *Quarterly Journal of Economics*, **117**, 1295–1328.

Mankiw, N. G., Reis, R., and Wolfers, J. (2003). Disagreement about inflation expectations. mimeo, National Bureau of Economic Research, Cambridge MA.

Mankiw, N. G., and Shapiro, M. D. (1986). News or noise: An analysis of GNP revisions. *Survey of Current Business (May 1986), US Department of Commerce, Bureau of Economic Analysis*, 20–25.

Mincer, J., and Zarnowitz, V. (1969). The evaluation of economic forecasts. In Mincer, J. (ed.), *Economic Forecasts and Expectations*, pp. 3–46. New York: National Bureau of Economic Research.

Patton, A. J., and Timmermann, A. (2010). Why do forecasters disagree? lessons from the term structure of cross-sectional dispersion. *Journal of Monetary Economics*, **57(7)**, 803–820.

Rich, R., and Tracy, J. (2010). The relationships among expected inflation, disagreement, and uncertainty: Evidence from matched point and density forecasts. *Review of Economics and Statistics*, **92(1)**, 200–207.

Rich, R. W., and Butler, J. S. (1998). Disagreement as a measure of uncertainty: A comment on Bomberger. *Journal of Money, Credit and Banking*, **30**, 411–419.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, **50**, 665–690.

Stock, J. H., Wright, J. H., and Yogo, M. (2002). A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business & Economic Statistics*, *20*(4), 518–529.

Woodford, M. (2002). Imperfect common knowledge and the effects of monetary policy. In Aghion, P., Frydman, R., Stiglitz, J., and Woodford, M. (eds.), *Knowledge, Information, and Expectations in Modern Macroeconomics: In honor of Edmund Phelps*, pp. 25–58: Princeton University Press.

Zarnowitz, V. (1969). The new ASA-NBER Survey of Forecasts by Economic Statisticians. *The American Statistician*, **23, No. 1**, 12–16.

Zarnowitz, V., and Lambros, L. A. (1987). Consensus and uncertainty in economic prediction. *Journal of Political Economy*, **95(3)**, 591–621.

5 Appendix A

Given the data generation process (DGP) defined by (1) to (3), and the information set $\mathcal{I}^t$, the optimal forecast of y_t^{t+2} is given by:

$$E\left(y_t^{t+2}|\mathcal{I}^t\right) = E\left(fy_{t-1}^{t+1} + v_t|\mathcal{I}^t\right) = fE\left(y_{t-1}^{t+1}|\mathcal{I}^t\right).$$

The minimum MSE predictor of $E\left(y_{t-1}^{t+1}|\mathcal{I}^t\right)$, given y_{t-1}^t, is:

$$E\left(y_{t-1}^{t+1}|\mathcal{I}^t\right) = fy_{t-2}^t + \gamma_H\left(y_{t-1}^t - h\left(y_{t-2}^{t-1} - y_{t-2}^t\right) - fy_{t-2}^t\right)$$

(see Kishor and Koenig (2012), for example). Using the shorthand notation that the forecast of the second estimate of y_t is denoted $\hat{y}_t^2$, and the first estimate by $\hat{y}_t^1$, we have:

$$\begin{aligned}
\hat{y}_t^2 &= fE\left(y_{t-1}^{t+1}|\mathcal{I}^t\right) \\
\hat{y}_t^1 &= \hat{y}_t^2 + h\left(y_{t-1}^t - E\left(y_{t-1}^{t+1}|\mathcal{I}^t\right)\right)
\end{aligned}$$

where $h\left(y_{t-1}^t - E\left(y_{t-1}^{t+1}|\mathcal{I}^t\right)\right) = E\left(\eta_t|\mathcal{I}^t\right)$.

Then we can derive the population values of β in (4) when f_t is either $\hat{y}_t^1$ or $\hat{y}_t^2$, when the model is estimated by TSLS (with suitable instruments, e.g., y_{t-1}^t, y_{t-2}^t and y_{t-2}^{t-1}).

When the LHS variable in (4) is $y_t^{t+1} - \hat{y}_t^1$, so that the first estimate is being targeted ($f_t = \hat{y}_t^1$), the estimate of β is:

$$\frac{Cov\left(y_t^{t+1} - \hat{y}_t^1, -E\left(\eta_t|\mathcal{I}^t\right)\right)}{Var\left(-E\left(\eta_t|\mathcal{I}^t\right)\right)}$$

where $-E\left(\eta_t|\mathcal{I}^t\right)$ is the first-stage predicted value, given the model of data revisions. Writing $y_t^{t+1} - \hat{y}_t^1$ as $\left(y_t^{t+2} - \hat{y}_t^2\right) + \left(\eta_t - E\left(\eta_t|\mathcal{I}^t\right)\right)$, the numerator is zero because $Cov\left(\left(y_t^{t+2} - \hat{y}_t^2\right), E\left(\eta_t|\mathcal{I}^t\right)\right) = 0$ (the noise η_t is orthogonal to y_t^{t+2}, so the predictable component of noise is orthogonal to the error in predicting y_t^{t+2}), and $Cov\left(\left(\eta_t - E\left(\eta_t|\mathcal{I}^t\right)\right), E\left(\eta_t|\mathcal{I}^t\right)\right) = 0$. Hence $\beta = 0$.

When the LHS variable in (4) is $y_t^{t+1} - \hat{y}_t^2$, i.e., $f_t = \hat{y}_t^2$, the estimate of β is:

$$\frac{Cov\left(y_t^{t+1} - \hat{y}_t^2, -E\left(\eta_t|\mathcal{I}^t\right)\right)}{Var\left(-E\left(\eta_t|\mathcal{I}^t\right)\right)}$$

Writing $y_t^{t+1} - \hat{y}_t^2$ as $(y_t^{t+2} - \hat{y}_t^2) + \eta_t$, then $Cov\left(y_t^{t+1} - \hat{y}_t^2, -E\left(\eta_t|\mathcal{I}^t\right)\right) = -Cov\left(\eta_t, E\left(\eta_t|\mathcal{I}^t\right)\right) < 0$, and $\beta \to -1$ as $\eta_t - E\left(\eta_t|\mathcal{I}^t\right) \to 0$, i.e., revisions become increasingly predictable